Uncovering Time-Specific Heterogeneity in Regression Discontinuity Designs

By: Mauricio Villamizar-Villegas
Yasin Kürşat Önder

Borradores de
ECONOMÍA

No. 1141
2020

Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia

Uncovering Time-Specific Heterogeneity in Regression Discontinuity Designs*

Mauricio Villamizar-Villegas† Yasin Kürşat Önder‡

Abstract

The literature that employs Regression Discontinuity Designs (RDD) typically stacks data across time periods and cutoff values. While practical, this procedure omits useful time heterogeneity. In this paper we decompose the RDD treatment effect into its weighted time-value parts. This analysis adds richness to the RDD estimand, where each time-specific component can be different and informative in a manner that is not expressed by the single cutoff or pooled regressions. To illustrate our methodology, we present two empirical examples: one using repeated cross-sectional data and another using time-series. Overall, we show a significant heterogeneity in both cutoff and time-specific effects. From a policy standpoint, this heterogeneity can pick up key differences in treatment across economically relevant episodes. Finally, we propose a new estimator that uses all observations from the original design and which captures the incremental effect of policy given a state variable. We show that this estimator is generally more precise compared to those that exclude observations exposed to other cutoffs or time periods. Our proposed framework is simple and easily replicable and can be applied to any RDD application that carries an explicitly traceable time dimension.

JEL Classification: C14, C31, C52

Keywords: Regression discontinuity, multiple cutoffs, time heterogeneity

*We especially thank David Phillips, Guido Kuersteiner, and Marinho Bertanha for valuable comments. We would also like to thank Maria Alejandra Ruiz for her excellent research assistance.

†Central Bank of Colombia; email: mvillavi@banrep.gov.co

‡Ghent University; email: kursat.onder@ugent.be

Heterogeneidad de tiempo en diseños de regresión discontinua

Mauricio Villamizar-Villegas Yasin Kürşat Önder

Abstract

La literatura que emplea diseños de regresión discontinua (RDD) generalmente agrupa observaciones a través del tiempo y a través de valores de corte. Si bien este método es práctico, puede omitir heterogeneidad útil de tiempo. En este documento descomponemos el efecto del tratamiento de RDD en sus partes ponderadas, relacionadas a cada valor temporal. De esta forma, nuestro análisis agrega riqueza al coeficiente estimado, donde cada componente específico del tiempo puede ser diferente e informativo, de una manera que no se expresa actualmente en las estimaciones de corte único o de cortes combinados. Para ilustrar nuestra metodología, presentamos dos ejemplos empíricos: uno usando datos de corte transversales repetidos y otro usando series de tiempo. En general, mostramos que existe una heterogeneidad significativa en los efectos de tiempo. Esta heterogeneidad puede generar diferencias relevantes en periodos económicos. Finalmente, proponemos un nuevo estimador que utiliza todas las observaciones del diseño original y que captura el efecto incremental de la política condicional a una variable de estado. Este estimador es generalmente más preciso en comparación con aquellos que excluyen observaciones expuestas a otros umbrales. Nuestra metodología es simple y fácilmente replicable y se puede aplicar a cualquier aplicación de RDD que tenga una dimensión rastreable de tiempo.

Clasificación JEL: C14, C31, C52

Palabras Clave: Regresión discontinua, múltiples valores de corte, heterogeneidad de tiempo

1 Introduction

In the standard Regression Discontinuity Design (**RDD**) setup, it is common to normalize (or at least de-mean) the running variable in order to pool observations together, regardless of whether they belong to different time periods or cutoffs. For instance, studies that evaluate incumbency advantages in electoral races will often combine constituencies that vary in the number of running candidates (Lee, 2008). In these cases, a voting margin share (relative to the runner-up) will most likely omit useful characteristics of each race, e.g. winning an election with a 10 percentage point margin in a 2-candidate race (say 55% - 45%) might be different than in a 3-candidate race (say 40% - 30% - 30%).

Moreover, conditions of the design can change over time, even when facing the same cutoff or number of candidates. For example, winning an election in times of economic crises can be different than in booms, especially if an incumbent is running for re-election. This applies to repeated cross-sectional data as well as time-series. Further, if time-specific policies, conditions, or individuals differ in systematic ways (and are correlated with the outcome or running variables), then the external validity of the design is compromised. More generally, the conventional normalizing-and-pooling approach –while consistent in most cases– will not precisely capture each individual time value effect.

To date, these issues have remained fairly unexplored. To the best of our knowledge, Cattaneo et al. (2016) is the first study to raise identification issues in a multi-cutoff setting, showing that the pooled RDD parameter identifies a weighted average of heterogeneous treatment effects. We also document Bertanha (2020) who proposes new estimation and inference methods for global average effects in settings with many thresholds. Using simulations and data from Romanian high-schools, the author uses variation in cutoff characteristics to predict average effects of policies that fall within the support of the variation of cutoff values and treatment doses. In turn, Cattaneo et al. (2020) use a design-based method to extrapolate multi-cutoff effects, with an identifying assumption that is analogous to the idea of common trends in a difference-in-difference estimation.[1]

In our investigation, we extend the multi-cutoff characterization to fit a time dimension, where heterogeneous treatment effects are accounted for at every time period. This extension

[1]Empirical examples of RDD with multiple cutoffs and treatments include the works of: Egger and Koethenbuerger (2010), Garibaldi et al. (2012), Pop-Eleches and Urquiola (2013a), Humlum et al. (2017), Francis-Tan and Tannuri-Pianto (2018), Zimmerman (2019), Önder and Shamsuddin (2019) and Fort et al. (2020).

adds richness to the RDD estimand, where each of its time-specific components can be different (and informative) in a manner that is not expressed by the single cutoff or pooled RDD regressions.

The main caveat, however, is the inherent tradeoff between richness and power, given that each time effect is even more localized and with a fewer number of observations. Put differently, the design not only narrows in on the vicinity of the running variable, but also on a given subperiod. To overcome this issue, we propose an estimator that uses all observations from the original design and which captures the incremental effect of policy given a state variable. Specifically, the state variable embedded in our non-parametric analysis is related to the design itself; directly if we consider multi-cutoff values, or indirectly if we consider different time periods (which can in turn be related to the different cutoffs). Hence, the use of this estimator in a multi-year and/or multi-cutoff setting constitutes an additional (and stand-alone) contribution of our work.

To illustrate our results, we replicate and extend two published empirical works. The first, based on Kuersteiner et al. (2018), evaluates a rule-based foreign exchange intervention mechanism enacted by the Colombian Central Bank, within a time-series framework. The design uses a cutoff rule with a sharp trigger: whenever the exchange rate exceeded a specific cutoff vis-a-vis its past moving average. Interestingly, the duration of the program lasted for almost a decade (2002-2012) and the cutoff value underwent several modifications. This allows us to exploit a multi-cutoff and multi-year analysis.

Our results applied to this study show that our proposed estimator is generally more precise (with narrower confidence intervals). Also, we display a significant heterogeneity in the cutoff-specific effects. And, since different cutoff values were enacted in mutually exclusive time periods, we also corroborate heterogeneity across cutoffs by evaluating year-specific effects. Finally, we show that our methodology can be used to answer other policy-oriented questions such as whether interventions are more effective in episodes of high exchange rate volatility. In this sense, similar to Vargas-Herrera and Villamizar-Villegas (2019) and for all time windows (post-intervention) considered, we find that interventions are more effective when volatility is high and less effective when low. The rationale according to Vargas-Herrera and Villamizar-Villegas, is that in the latter case agents are willing to bet against the central bank, while in the former case intervention faces a weaker countervailing force from speculators.

The second exercise, based on Lee et al. (2004), estimates the effect of a candidate's electoral strength on subsequent roll-call voting records, using repeated cross-sectional data. In essence, the authors evaluate whether electoral competition induces candidates to compromise their political views and move towards a more moderate standing. That is, they test whether voters *affect* candidates' policy choices or if they merely *elect* them. The study uses voting record data from the U.S. House of Representatives during the period of 1946-1995, which allows us to evaluate time-specific effects for a period that spans for five decades.

Our results show, once more, a gain in statistical power of the incremental estimator. We then evaluate whether the estimated effects vary depending on whether the economy is under or over performing. After all, incumbency advantage can be amplified in times of economic booms and reduced in crises. In periods of economic upturns, we find a positive "elect" component and a generally non-significant "affect" component. These results are consistent with Lee et al. (2004), although our estimates are lesser in magnitude. However, results are much different in episodes of economic downturns, where we find a positive and significant "affect" component. This implies that, in economic crises, voters *affect* and *elect* policies. This result partially vindicates Downs' (1957) paradigm of the median voter, where political competition leads to a policy convergence among candidates.

Besides our two empirical examples, which highlight that a researcher misses important information without considering the time dimension, characterizing time heterogeneity can have useful applications in a wide range of economic issues. As such, our framework can be extended to almost all studies that use repeated cross-sectional or time-series data. To name a few, one can potentially exploit the time heterogeneity found in: Auffhammer and Kellogg (2011) who evaluate the impact of gasoline content regulation on air quality, Ito (2015) who explores whether an electricity rebate program enacted in California leads to energy conservation, Chay and Greenstone (2003) who investigate the impact of the 1970 Clean Air Act Amendments on infant mortality, Pop-Eleches and Urquiola (2013b) who study the effects of attending high achievement schools on students' academic performance, Zimmerman (2019) who investigates whether top colleges alleviate the barriers for historically disadvantaged groups to reach top positions in the economy, Fort et al. (2020) who report significant cognitive and non-cognitive costs of day care, and Chay et al. (2005) who study whether better infrastructure and resource allocation in schools lead to better test scores. For a more thorough compilation of empirical findings using RDDs we refer readers to the surveys of Lee and Lemieux (2010) and Villamizar-Villegas et al. (2020).

2 Methodology

In a sharp multi-cutoff RDD setting, the probability of facing a particular cutoff can be expressed as a function of individuals' characteristics. In this section we allow that cutoff value to be time dependent. Formally, the assignment of treatment, D_{it}, is completely determined by a cutoff-rule based on a continuous running variable, X_{it}, as follows:

$$D_{it} = \mathbf{1}\left\{X_{it} \geq c_{it}\right\} \tag{1}$$

where $\mathbf{1}$ denotes an indicator function and c_{it} is the time-dependent threshold below which treatment is denied. We henceforth assume that the cutoff has finite support and that the running variable is observable and continuous (standard assumptions in the RDD setting). Note that the time-dependent multi-cutoff can be stated as a random variable C_{it} that each individual i faces at time t, with a support defined in JxT, so that:

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1T} \\ c_{21} & c_{22} & \cdots & c_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ c_{J1} & c_{J2} & \cdots & c_{JT} \end{bmatrix}.$$

As a result, individuals may face J possible cutoff values in each of the T time periods. Notice that the added time and multi-cutoff dimensions exponentially increase the number of potential outcomes. That is, while in the single-cutoff framework an individual has only two potential outcomes based on a dichotomous treatment (Y_{1i} if exposed to treatment and Y_{0i} if not exposed), in this case each individual has (2) x (J) x (T) potential outcomes, so that $Y_{dit}(C_{it}) = \sum_{c \in C} \mathbf{1}\left\{C_{it} = c\right\} Y_{dit}(c)$, for $d = 0, 1$.

It follows that the pooled time-specific multi-cutoff average treatment effect, τ, is obtained by comparing the conditional mean of an outcome variable at the limit on either side of the *various* discontinuity points:

$$\begin{aligned} \tau &= E\left(Y_{1it}(C_{it}) - Y_{0it}(C_{it}) \mid X_{it} = C_{it}\right) \\ &= \lim_{\epsilon \to 0^+} E\left(Y_{it}(C_{it}) \mid X_{it} = C_{it} + \epsilon\right) - \lim_{\epsilon \to 0^+} E\left(Y_{it}(C_{it}) \mid X_{it} = C_{it} - \epsilon\right), \end{aligned} \tag{2}$$

where the final equality holds as long as the conditional distributions of potential outcomes, $\Pr\left(Y_{1it} \leq y \mid X_{it} = x\right)$ and $\Pr\left(Y_{0it} \leq y \mid X_{it} = x\right)$, are continuous in x at $x = c$ for all $c \in C$.

Essentially, τ averages effects across time periods and cutoffs. This is perhaps more illustrative in the following equation:

$$\tau = \sum_{c \in C} E\left(Y_{1it}(c) - Y_{0it}(c) \mid X_{it} = c, C_{it} = c\right) \left(\frac{f_{X|C}(c \mid c) Pr(C_{it} = c)}{\sum_{c \in C} f_{X|C}(c \mid c) Pr(C_{it} = c)} \right) \tag{3}$$

where $f_{X|C}(x \mid c)$ is the conditional density of X_{it}. The last term of equation 3 denotes the weight given to the overall effect, which increases with the probability of observing a realization of each cutoff in a given time period. Consequently, τ extracts more information from the design if a particular sub-sample contains more observations closer to the cutoff.

The main advantage of decomposing the treatment effect into its weighted parts is that a researcher might be interested in a particular time effect. For instance, in the electoral example found in Lee et al. (2004), it might be useful to differentiate the effects of barely winning an election for specific cases (e.g. when the economy is under or over performing). Alternatively, in the foreign exchange intervention example found in Kuersteiner et al. (2018), it can be useful for central banks to know whether the impact of an intervention mechanism increases depending on whether exchange rate volatility is high or low. Ultimately, when enough time periods or conditions are evaluated, a treatment time curve can be obtained.

The main caveat however, is the inherent tradeoff between richness and power, given that each time effect is even more localized and with a fewer number of observations. As stated in Cattaneo et al. (2016), for every cutoff $c \in C$, point estimators are constructed by "*keeping only the observations exposed to the cutoff c and then employing directly local polynomial methods treating the cutoff c as the single cutoff in this subsample.*" Inevitably, this procedure reduces the number of observations and consequently, the non-rejection of the null hypothesis (H_0 : no treatment effect) can be partly attributable to a decrease in statistical power.

To overcome this problem, we propose an estimator that uses all observations from the original design and which captures the incremental effect of policy given a state variable. Intuitively, we include an interaction term –between treatment and a particular period and/or cutoff– within the standard pooled regression.[2]

[2]We refer readers to the the works of Angrist and Rokkanen (2015) and Bertanha and Imbens (2020) for an in-depth review of identifying assumptions for observations away from the threshold, which also apply to our framework.

To better illustrate, consider an example with cumulative cutoffs where different treatments are given for different values of the running variable. Such is the case of a central bank that intervenes in the foreign exchange market by purchasing foreign currency based on the exchange rate: suppose a small intervention value if the exchange rate appreciates by more than 2%, a medium-sized value if it appreciates by more than 4%, and a high value if it appreciates by more than 5%. In this case, each value of the exchange rate is exposed to only one of the three cutoffs.

Now suppose that we are interested in the effects of crossing the 2% threshold (i.e. small purchases of foreign currency). With a sufficiently wide bandwidth, the design runs the risk of including values above 4% which correspond to a larger size intervention. As such, the recommendation in Cattaneo et al. (2016) is to "*only include observations whose running variable is not smaller than the cutoff immediately before and no larger than the cutoff immediately after*." This implies considering exchange rate appreciation values below 4%. Notwithstanding, the design now ignores useful information (which can reduce variability of the dependent variable) above and including the 4% threshold.

Our proposed estimator overcomes this issue by including observations exposed to other cutoffs, while focusing on the cutoff of interest. For ease in exposition, first consider a cross-sectional setup with J different cutoffs, so that $C = \{c_1, c_2, \cdots, c_J\}$. Let $D_i = \mathbf{1}\{X_i \geq c_i\}$ be the assignment of treatment pertaining to each cutoff and which is different from the assignment variable in the pooling approach $\tilde{D}_i = \mathbf{1}\{X_i \geq c\}$. Similarly, denote $\tilde{X}_i = X_i - c$ as the centered running variable that pools all observations together (just like in the conventional single cutoff framework), whereas $X_i = X_i - c_i$ only considers observations exposed to the cutoff c_i.

In this case, a researcher can obtain the incremental effect a particular cutoff (c_i) on any outcome variable y_t, by estimating the following:

$$\arg\min \sum_{i=1}^{J} \left(y_i - \alpha - \beta_1 D_i - \beta_2 \tilde{X}_i - \beta_3 \tilde{X}_i D_i \right)^2 K\left(\frac{\tilde{X}_i}{h}\right) \qquad \text{for } i = 1, 2, ..., J \tag{4}$$

where $K(\cdot)$ is a kernel function with bandwidth h and which assigns weights to each observation based on the distance between X_i and c. Intuitively, these weights can be interpreted as the relative ex-ante probability of the running variable falling within the immediate neighborhood of the threshold (Lee and Lemieux, 2010). Also, in this framework β_2 and β_3 can take the form of polynomial functions of the running variable. In Appendix A we provide further intuition of this specification.

It follows that β_1 captures the incremental effect of each cutoff c_i over the average treatment effect, with the advantage of using the same sample size as the normalizing-and-pooling approach. Notice that in the framework presented in Cattaneo et al. (2016) equation 4 would have X_i instead of $\tilde{X}_i$, meaning that only observations exposed to the cutoff c_i would be considered. Alternatively, in a normalizing-and pooling approach, the term D_i would be replaced with $\tilde{D}_i$. Analogously, in a time-series setting, the following regression can be estimated:

$$\arg\min \sum_{t=1}^{T} \left(y_t - a - b_1 D_t - b_2 \hat{X}_t - \beta_3 \hat{X}_t D_t \right)^2 K\left(\frac{\hat{X}_t}{h} \right) \qquad \text{for } t = 1, 2, ..., T \tag{5}$$

where $D_t = \mathbf{1}\{X_t \geq c_t\}$ and $\tilde{X}_t = X_t - c$. Even more generally, to estimate the incremental effect of a given time period, one can simply replace $D_t$by a time dummy variable of interest. The coefficient β_1 would then evaluate the incremental effect of that specific time-value effect.

3 Empirical Examples

3.1 Effectiveness of Foreign Exchange Intervention

We begin by illustrating the design in Kuersteiner et al. (2018). During 2002-2012, the Central Bank of Colombia enacted a rule-based mechanism aimed at curbing exchange rate volatility. Specifically, auctions (consisting of FX options) were triggered whenever the Peso-Dollar exchange rate (COP:USD) vis-à-vis its last 20-day moving average exceeded an established cutoff. On average, the amount of foreign currency announced at each auction was \$180 million dollars, representing roughly 25% of the daily FX market turnover.

The cutoff value was initially set at ±4% (put options were triggered below -4% and call options were triggered above 4%), but was later modified to ±2% from December 2005 until June 2008, and to ±5% from October 2008 until October 2009. This allows us to exploit both a multi-cutoff and multi-year analysis.

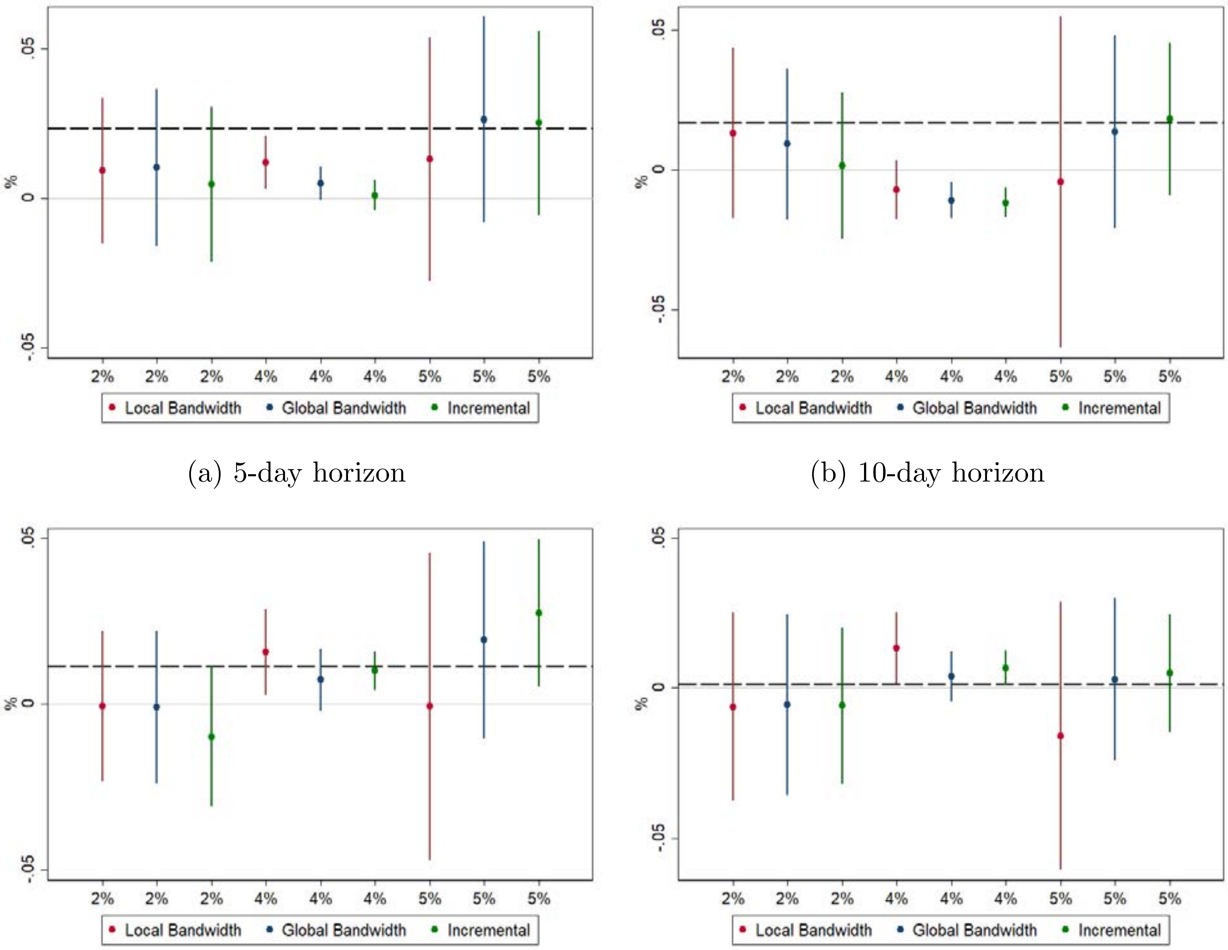
We present three estimators for each cutoff. The first is a local bandwidth estimator that follows the methodology in Cattaneo et al. (2016) by restricting the sample to the time period where each cutoff value was enacted. For comparability purposes, we use a triangular kernel with an optimal bandwidth choice resulting from the cross-validation procedure described in

Imbens and Kalyanaraman (2012). The second estimator (global bandwidth) also restricts the sample to the time period where each cutoff was enacted, but where we manually impute a global bandwidth, as if we had instead pooled all observations together. Note that the sole purpose of this estimator is to graphically show the transition from the individual component of each cutoff to the benchmark case presented in Kuersteiner et al. and which is depicted as a dashed horizontal line in each figure. Finally, the third estimator (incremental) uses all observations from the original design; it is nonetheless localized in the sense that the effect of the running variable is evaluated close to the cutoff c_j, but includes observations exposed to all the different cutoffs, as shown in equation 5.

Figure 1 depicts our results across cutoffs which, by design, coincide with the different and mutually exclusive time periods. Panels 1a - 1d show the effects of the central bank auction on the exchange rate at horizons of 5, 10, 15, and 20 days. As expected, the incremental estimator is generally more precise (with narrower confidence intervals). Also, there is significant heterogeneity in the cutoff-specific effects, especially at horizons of 5 and 10 days. Overall, the effects tend to be greater for when the cutoff was placed at 5% which coincides with the latest years of the sample (2008 - 2009) in which interventions were triggered. That is, while the rule-based mechanism was in place until February 2012, the rule was last triggered in July 2009. We corroborate this in Figure B1 of Appendix B where we plot the effects by year and which shows a noticeable increase during 2008 and 2009.

Perhaps more useful to policymakers, Figure 2 compares the effects of FX intervention in episodes of high exchange rate volatility with episodes of low volatility. Similar to Vargas-Herrera and Villamizar-Villegas (2019), we define periods of high volatility as those in which exchange rate values (in a 20-day moving window) exceeded a one standard deviation. For all time windows (post-intervention) considered, we confirm the findings in Vargas-Herrera and Villamizar-Villegas: intervention is more effective when volatility is high and less effective when low. The rationale according to the authors, is that in the latter case agents are willing to bet against the central bank, while in the former case intervention faces a weaker countervailing force from speculators.

Figure 1: Exchange rate effects across the 2%, 4%, and 5% cutoffs



(a) 5-day horizon

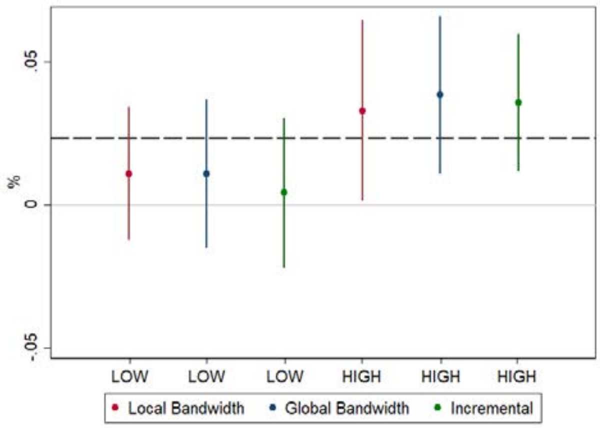
(b) 10-day horizon

(c) 15-day horizon

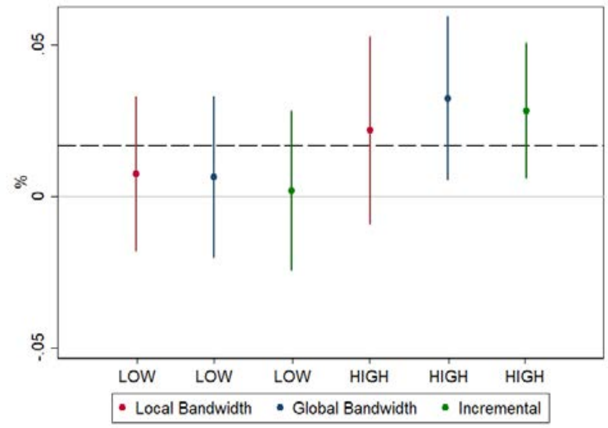
(d) 20-day horizon

The figure shows three separate regression discontinuity estimates: Local, Global, and Incremental, for different time horizons. The response of the outcome variable is expressed in exchange rate changes (in %) and the impulse is a rule-based FX intervention (purchases) of $180 million dollars, as presented in Kuersteiner et al. (2018). We use a triangular kernel and optimal bandwidths from the cross-validation procedure in Imbens and Kalyanaraman (2012). Confidence intervals based on heteroskedasticity-robust standard errors are set at a 5% significance level.

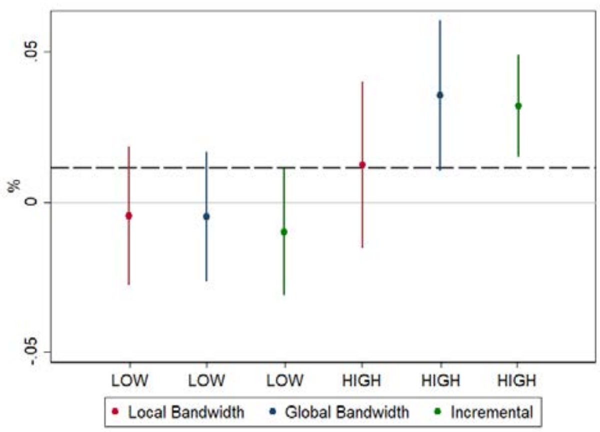
Figure 2: Exchange rate effects in episodes of low and high exchange rate volatility



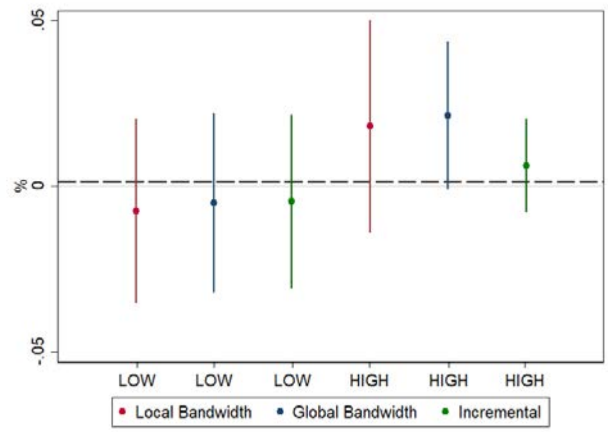
(a) 5-day horizon



(b) 10-day horizon



(c) 15-day horizon



(d) 20-day horizon

The figure shows three separate regression discontinuity estimates: Local, Global, and Incremental, for different time horizons. The response of the outcome variable is expressed in exchange rate changes (in %) and the impulse is a rule-based FX intervention (purchases) of $180 million dollars, as presented in Kuersteiner et al. (2018). Periods of high exchange volatility are defined as those in which exchange rate values (in a 20-day moving window) exceeded a one standard deviation. We use a triangular kernel and optimal bandwidths from the cross-validation procedure in Imbens and Kalyanaraman (2012). Confidence intervals based on heteroskedasticity-robust standard errors are set at a 5% significance level.

3.2 Electoral Strength on Roll-Call Voting Records

Lee et al. (2004) evaluate the effect of a candidate's electoral strength on subsequent roll-call voting records. Essentially, the authors explore whether voters can affect candidates' policy choices or if they merely elect them. In the former, candidates compete for votes by compromising their views towards a more moderate standing, while in the latter, the legislator's voting behavior is independent from any electoral support.

In the design, random variation in electoral strength is attained by focusing on US House elections decided by narrow margins during the period of 1946-1995. So, if for example a Democratic candidate barely wins a seat in one election, then the Democratic party holds a relatively exogenous incumbency advantage in the next election. Hence, under the premise that voters actually do affect policies, it follows that "*when electoral strength is high, a candidate can afford to vote in a relatively more partisan way if he is elected; a weaker candidate would be forced to choose a more moderate policy.*"[3]

To measure the degree of political compromise, the authors use the voting score by the Americans for Democratic Action (ADA), where a high score (index from 0-100) denotes a more liberal record. From a Democratic (Republican) perspective, a decrease (increase) in the ADA score reflects a compromise in ideology. As it turns out, Lee et al. find that the degree of electoral strength has no effect on a legislator's voting behavior. In other words, candidates with weak (strong) electoral support do not adopt more (less) moderate positions.

In this investigation we take on a fresh new angle by evaluating whether this result holds depending on whether the economy is under or over performing. After all, incumbency advantage can be amplified in times of economic booms and reduced in crises. Technically speaking, this does not necessarily imply that electoral support is lost during a crisis, but rather, that politicians are more likely to be "*stress tested*" (by voters) during these periods.

We begin by corroborating the authors' findings for every decade taken separately. In Table 1 we report three estimators for each effect: (i) the result reported by the authors, (ii) a local estimator that restricts the sample to each decade, and (iii) our incremental estimator that evaluates the effect of each decade, but which includes observations from the entire sample. In principle, estimators (i) and (ii) should be identical. This is generally the case given that the minor differences are attributable to decimal rounding and reporting.

[3]Lee et al. (2004), page 809.

However, for the "elect" and "affect" components (columns 4 and 5) we follow the procedure in Button (2015) to estimate the standard errors by using a clustered bootstrap (clustering over district-by-decade), since it is unclear how Lee et al. (2004) estimate the standard errors for these components. Similar to the previous example of Section 3.1, we highlight the gain in statistical power of the incremental estimator. In terms of the number of observations, it uses 915 observations, while estimators (i) and (ii) use 322 in 1946–1958; 245 in 1960–1968; 183 in 1970–1978; and 164 in 1980–1996.

The main findings of Table 1 are as follows: Column (1) shows results for the total effect. They capture the liberal ADA score difference (in period $t+1$) between a Democratic winner whose previous congressional seat (in period t) was occupied by a barely chosen Democrat and a Democratic winner whose previous congressional seat was occupied by a barely chosen Republican (and vice-versa). As shown, results for the incremental estimator are very different from the local estimators. In some decades, the magnitude is greatly reduced: by close to 80% in the 1960's and by half in the 1980's. In relative terms, the effects in the 1970's change rank from smallest (with estimators *i* and *ii*) to the second largest (with estimator *iii*).

Column (4) shows results for the "elect" component. Intuitively, it captures how the ADA score is higher (more liberal) simply because the winner is more likely to be a Democrat. It is constructed by multiplying column (2) which is the effect of the party's affiliation with column (3) which is the effect of the party's initial win on winning the next election. As observed, results for the incremental estimator are much smaller in magnitude, but still statistically significant.

Finally, Column (5) shows results for the "affect" component, computed as the total effect (column 1) minus the "elect" component (column 4). It reflects how candidates respond to a change in the probability of winning an election. According to the authors' findings, this effect is not significantly different from zero. However, our local estimator for the decade of 1970-1978 shows a negative and significant impact, while the incremental estimator shows a positive and significant effect for the decade of 1980-1996.

Table 1: Effects of electoral strength on roll-call voting records by decade

Variable	(1) Total effect γ ADA_{t+1}	(2) π_1 ADA_t	(3) $\left(P^D_{t+1} - P^R_{t+1}\right)$ DEM_{t+1}	(4) Elect component $\pi_1\left[P^D_{t+1} - P^R_{t+1}\right]$ col.(2)*col.(3)	(5) Affect component $\pi_0\left[P^{*D}_{t+1} - P^{*R}_{t+1}\right]$ col.(1)-col.(4)
1946-1958					
Authors' calculation	14.2***	41.7***	0.41***	17.0***	-2.8
	(3.2)	(2.3)	(0.05)	(4.8)	(4.0)
Local	14.24***	41.80***	0.42***	17.34***	-3.10
	(3.22)	(2.33)	(0.05)	(2.46)	(2.50)
Incremental	11.32***	25.92***	0.33***	8.63***	2.69
	(2.68)	(1.87)	(0.04)	(1.70)	(2.61)
1960-1968					
Authors' calculation	23.5***	49.5***	0.51***	25.2***	-1.7
	(3.5)	(2.7)	(0.05)	(4.9)	(4.1)
Local	23.59***	49.51***	0.51***	25.26***	-1.67
	(3.56)	(2.72)	(0.05)	(4.02)	(2.58)
Incremental	5.17*	26.90***	0.16***	4.24**	0.93
	(2.86)	(2.66)	(0.04)	(1.69)	(2.72)
1970-1978					
Authors' calculation	11.5**	46.6***	0.40***	18.6***	-7.1
	(4.7)	(3.1)	(0.06)	(5.1)	(5.1)
Local	11.55**	46.63***	0.40***	18.83***	-7.28*
	(4.73)	(3.19)	(0.068)	(4.15)	(4.42)
Incremental	13.34***	30.30***	0.30***	9.06***	4.28
	(3.35)	(2.77)	(0.047)	(2.41)	(3.07)
1980-1996					
Authors' calculation	46.8***	56.6***	0.76***	43.0***	3.8
	(3.7)	(2.8)	(0.05)	(4.9)	(4.5)
Local	46.84***	56.70***	0.77***	43.47***	3.37
	(3.73)	(2.90)	(0.05)	(5.52)	(2.20)
Incremental	23.80***	33.01***	0.362***	11.94***	11.87***
	(3.28)	(2.55)	(0.05)	(2.87)	(2.26)

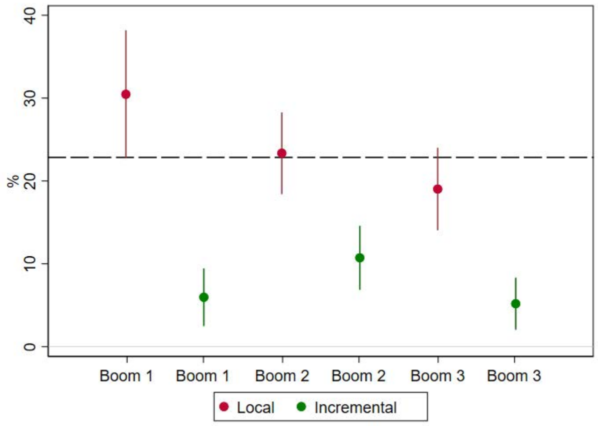
Authors' calculations. The table reports three separate regression discontinuity estimates: one used by Lee et al. (2004), a Local, and an Incremental estimator, for different decades. The unit of observation corresponds to a district-congressional session. ADA is the voting score by the Americans for Democratic Action, where a high score (index from 0-100) denotes a more liberal record. Heteroskedasticity-robust standard errors are in parentheses. The sample includes observations where the Democratic vote share is between 48-52% percent. Time t and $t+1$ denote congressional sessions. The incremental estimator uses 915 observations, while the other estimators use 322 in 1946–1958; 245 in 1960–1968; 183 in 1970–1978; and 164 in 1980–1996.

We further evaluate the "elect" and "affect" components in periods of economic booms. For robustness, we consider three definitions of booms: (i) episodes marked by the media as sustained periods of economic growth and which include: 1950-1951 (the "Golden Age"), 1962-1967 (tax cuts, government spending, and Vietnam war), and 1984-1989 (tax cuts and low interest rates),[4] (ii) periods with a positive output growth, and (iii) periods with a positive GDP gap, as calculated by the Federal Reserve Bank of St. Louis (FRED). Our results are depicted in Figures 3a and 3b and show a positive "elect" component (although lesser in magnitude for the incremental estimator) and a generally non significant "affect" component. Consequently, in times of economic upturns, our results are consistent to those of Lee et al. (2004).

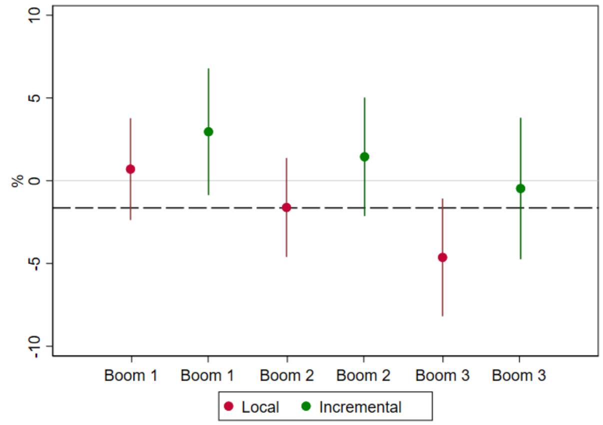
However, results are much different for economic downturns. Again we consider three definitions of crises: (i) episodes marked by the media and which include: 1973-1975 (the "Nixon Shock" which references the oil crisis and the demise of the Bretton Woods System, among others), 1980-1982 (the energy crsis), and 1990 (oil crisis and contractionary monetary policies),[5] (ii) periods with a negative output growth, and (iii) periods with a negative GDP gap. Results in Figure 3d now show a positive and significant "affect" component captured only by the incremental estimator. This implies that in economic crises, voters actually "affect" and "elect" policies. In some sense, this result partially revindicates Downs (1957) paradigm of the median voter, where political competition leads to a policy convergence among candidates.[6]

[4]See for example a related article from CNN.
[5]See for example a related article from BBC.
[6]Albouy (2011) arrives to a similar result by adjusting for the fact that junior members of Congress have more extreme views than senior members.

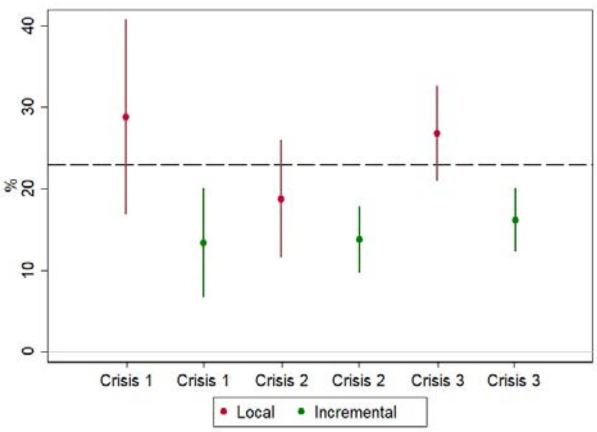
Figure 3: Effects of electoral Strength on voting records



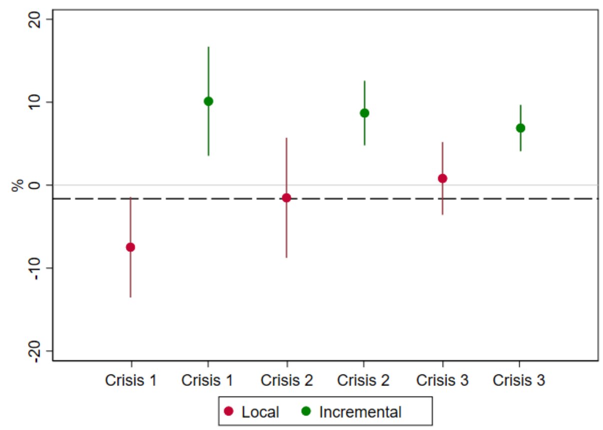
(a) *Elect* component in booms



(b) *Affect* component in booms



(c) *Elect* component in crises



(d) *Affect* component in crises

The figure shows two separate regression discontinuity estimates: Local and Incremental, for different measures of economic booms (panels *a* and *b*) and crises (panels *c* and *d*). The *elect* component is constructed by multiplying the effect of the party's affiliation (column 2 of Table 1) with the effect of the party's initial win on winning the next election (column 3 of Table 1). The *affect* component is computed as the total effect (column 1 of Table 1) minus the *elect* component (column 4 of Table 1). Economic booms (crises) are defined as: (1) episodes marked by the media, (2) periods with a positive (negative) GDP growth, and (3) periods with a positive (negative) GDP gap. Confidence intervals based on heteroskedasticity-robust standard errors are set at a 5% significance level.

4 Conclusion

The growing literature that employs Regression Discontinuity Designs (RDDs) generally normalizes the running variable in order to pool observations together, regardless of whether they belong to different time periods or cutoff values. While practical, this procedure omits useful time heterogeneity within each cutoff.

In this paper we decompose the treatment effect into its weighted time-value parts. This extension adds richness to the RDD estimand, where each time-specific component can be different (and informative) in a manner that is not expressed by the single cutoff or pooled RDD regressions. From a policy standpoint, this heterogeneity can pick up key differences in treatment effects across economically relevant episodes.

As a result, characterizing time heterogeneity can have useful applications in a wide range of economic issues. The main caveat however, is the inherent tradeoff between richness and statistical power, given that each time effect is even more localized. To overcome this problem, we propose an estimator that uses all observations from the original design and which captures the incremental effect of policy given a state variable.

We illustrate our framework by replicating and extending two published empirical studies. Importantly, we show that our proposed estimator is generally more precise (with narrower confidence intervals). Also, we show a significant heterogeneity in both cutoff-specific and time-specific effects. Hence, we argue that useful heterogeneity is usually left out of any RDD analysis.

Our proposed framework is simple and easily replicable. It can be applied to almost all RDD applications that have an explicitly traceable time dimension.

5 Bibliography

Albouy, D. (2011): "Do voters affect or elect policies? A new perspective, with evidence from the U.S. Senate." *Electoral Studies*, 30, 162 – 173.

Angrist, J. D. and M. Rokkanen (2015): "Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away From the Cutoff," *Journal of the American Statistical Association*, 110, 1331–1344.

Auffhammer, M. and R. Kellogg (2011): "Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality," *American Economic Review*, 101, 2687–2722.

Bertanha, M. (2020): "Regression discontinuity design with many thresholds," *Journal of Econometrics*, 218, 216–241.

Bertanha, M. and G. W. Imbens (2020): "External Validity in Fuzzy Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 38, 593–612.

Button, P. (2015): "A Replication of 'Do Voters Affect or Elect Policies? Evidence from the U.S. House' (The Quarterly Journal of Economics, 2004)," Working Papers 1518, Tulane University, Department of Economics.

Cattaneo, M. D., L. Keele, R. Titiunik, and G. Vazquez-Bare (2020): "Extrapolating Treatment Effects in Multi-Cutoff Regression Discontinuity Designs," Forthcoming.

Cattaneo, M. D., R. Titiunik, G. Vazquez-Bare, and L. Keele (2016): "Interpreting regression discontinuity designs with multiple cutoffs," *The Journal of Politics*, 78, 1229–1248.

Chay, K. Y. and M. Greenstone (2003): "AIR QUALITY, INFANT MORTALITY, AND THE CLEAN AIR ACT OF 1970," NBER Working Paper 10053.

Chay, K. Y., P. J. McEwan, and M. Urquiola (2005): "The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools," *American Economic Review*, 95, 1237–1258.

Downs, A. (1957): "An Economic Theory of Political Action in a Democracy," *Journal of Political Economy*, 65, 135–150.

Egger, P. and M. Koethenbuerger (2010): "Government Spending and Legislative Organization: Quasi-experimental Evidence from Germany," *American Economic Journal: Applied Economics*, 2, 200–212.

Fort, M., A. Ichino, and G. Zanella (2020): "Cognitive and Noncognitive Costs of Day Care at Age 0–2 for Children in Advantaged Families," *Journal of Political Economy*, 128, 158–205.

Francis-Tan, A. and M. Tannuri-Pianto (2018): "Black Movement: Using discontinuities in admissions to study the effects of college quality and affirmative action," *Journal of Development Economics*, 135, 97–116.

Garibaldi, P., F. Giavazzi, A. Ichino, and E. Rettore (2012): "College Cost and Time to Complete a Degree: Evidence from Tuition Discontinuities," *The Review of Economics and Statistics*, 94, 699–711.

Humlum, M. K., J. H. Kristoffersen, and R. Vejlin (2017): "College admissions decisions, educational outcomes, and family formation," *Labour Economics*, 48, 215–230.

Imbens, G. and K. Kalyanaraman (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79, 933–959.

Ito, K. (2015): "Asymmetric Incentives in Subsidies: Evidence from a Large-Scale Electricity Rebate Program," *American Economic Journal: Economic Policy*, 7, 209–237.

Kuersteiner, G. M., D. C. Phillips, and M. Villamizar-Villegas (2018): "Effective sterilized foreign exchange intervention? Evidence from a rule-based policy," *Journal of International Economics*, 113, 118–138.

Lee, D. S. (2008): "Randomized experiments from non-random selection in US House elections," *Journal of Econometrics*, 142, 675–697.

Lee, D. S. and T. Lemieux (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48, 281–355.

Lee, D. S., E. Moretti, and M. J. Butler (2004): "Do voters affect or elect policies? Evidence from the US House," *The Quarterly Journal of Economics*, 119, 807–859.

Önder, Y. K. and M. Shamsuddin (2019): "Heterogeneous Treatment Under Regression Discontinuity Design: Application to Female High School Enrolment," *Oxford Bulletin of Economics and Statistics*, 81, 744–767.

Pop-Eleches, C. and M. Urquiola (2013a): "Going to a Better School: Effects and Behavioral Responses," *American Economic Review*, 103, 1289–1324.

——— (2013b): "Going to a Better School: Effects and Behavioral Responses," *American Economic Review*, 103, 1289–1324.

Vargas-Herrera, H. and M. Villamizar-Villegas (2019): "Effectiveness of FX Intervention and the Flimsiness of Exchange rate Expectations," Borradores de Economia 1070, Banco de la Republica de Colombia.

Villamizar-Villegas, M., F. A. Pinzón-Puerto, and M. A. Ruiz-Sánchez (2020): "A Comprehensive History of Regression Discontinuity Designs: An Empirical Survey of the last 60 Years," Borradores de Economia 1112, Banco de la Republica de Colombia.

Zimmerman, S. D. (2019): "Elite Colleges and Upward Mobility to Top Jobs and Top Incomes," *American Economic Review*, 109, 1–47.

Appendix A Incremental Effect Specification

We briefly provide some intuition behind equation 4. For simplicity, consider the following non-localized linear regression:

$$y_i = \alpha + \beta_1 D_i + \beta_2 X_i + \beta_3 F_i + \beta_4 D_i X_i + \beta_5 D_i F_i + \epsilon_i \tag{A1}$$

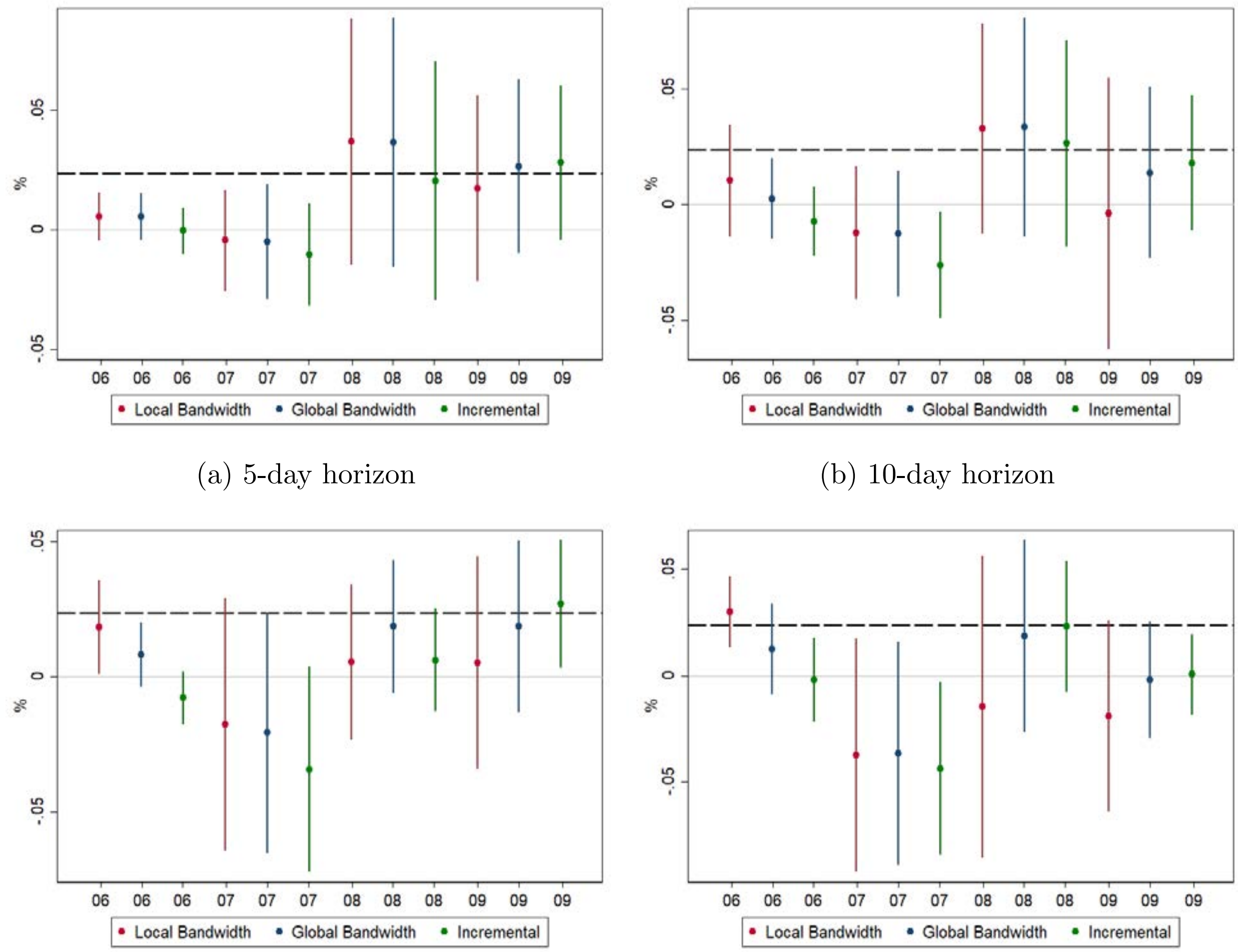
where D_i is the assignment of treatment which is based on values of the running variable X_i. Under certain exogeneity conditions, the coefficient β_1 would capture the average treatment effect, and thus, β_5 would capture the incremental effect of treatment for a given value of F_i. Now assume that F_i is actually a subset of D_i, i.e. a much stricter assignment of treatment. In this case, $F_i D_i = F_i$. It follows that the treatment effect of the stricter policy can also be estimated by b_1 in the following regression:

$$y_i = \alpha + b_1 F_i + b_2 X_i + b_3 F_i X_i + \eta_i \tag{A2}$$

which, in a localized framework, is equivalent to that of equation 4.

Appendix B Year Specific Effects

Figure B1: Exchange rate effects across years



(a) 5-day horizon

(b) 10-day horizon

(c) 15-day horizon

(d) 20-day horizon

The figure shows three separate regression discontinuity estimates: Local, Global, and Incremental, for different time horizons. The response of the outcome variable is expressed in exchange rate changes (in %) and the impulse is a rule-based FX intervention (purchases) of $180 million dollars, as presented in Kuersteiner et al. (2018). We use a triangular kernel and optimal bandwidths from the cross-validation procedure in Imbens and Kalyanaraman (2012). Confidence intervals based on heteroskedasticity-robust standard errors are set at a 5% significance level.


BANCO DE LA REPÚBLICA
COLOMBIA